BMJ 2014;349:g6501 doi: 10.1136/bmj.g6501 (Published 21 November 2014)

RESEARCH

Selective reporting bias of harm outcomes within studies: findings from a cohort of systematic reviews

Pooja Saini *research associate*[1], Yoon K Loke *professor*[2], Carrol Gamble *professor*[3], Douglas G Altman *professor*[4], Paula R Williamson *professor*[3], Jamie J Kirkham *lecturer*[3]

[1]Department of Public Health and Policy, University of Liverpool, Liverpool, UK; [2]Norwich Medical School, University of East Anglia, Norwich, UK; [3]Department of Biostatistics, University of Liverpool, Liverpool, L69 3GA, UK; [4]Centre for Statistics in Medicine, University of Oxford, Oxford, UK

Abstract

Objective To determine the extent and nature of selective non-reporting of harm outcomes in clinical studies that were eligible for inclusion in a cohort of systematic reviews.

Design Cohort study of systematic reviews from two databases.

Setting Outcome reporting bias in trials for harm outcomes (ORBIT II) in systematic reviews from the Cochrane Library and a separate cohort of systematic reviews of adverse events.

Participants 92 systematic reviews of randomised controlled trials and non-randomised studies published in the Cochrane Library between issue 9, 2012 and issue 2, 2013 (Cochrane cohort) and 230 systematic reviews published between 1 January 2007 and 31 December 2011 in other publications, synthesising data on harm outcomes (adverse event cohort).

Methods A 13 point classification system for missing outcome data on harm was developed and applied to the studies.

Results 86% (79/92) of reviews in the Cochrane cohort did not include full data from the main harm outcome of interest of each review for all of the eligible studies included within that review; 76% (173/230) for the adverse event cohort. Overall, the single primary harm outcome was inadequately reported in 76% (705/931) of the studies included in the 92 reviews from the Cochrane cohort and not reported in 47% (4159/8837) of the 230 reviews in the adverse event cohort. In a sample of primary studies not reporting on the single primary harm outcome in the review, scrutiny of the study publication revealed that outcome reporting bias was suspected in nearly two thirds (63%, 248/393).

Conclusions The number of reviews suspected of outcome reporting bias as a result of missing or partially reported harm related outcomes from at least one eligible study is high. The declaration of important harms and the quality of the reporting of harm outcomes must be improved in both primary studies and systematic reviews.

Introduction

> "When we looked at that data, it actually showed an increase in harm amongst those who got the active treatment, and we ditched it because we weren't expecting it and we were concerned that the presentation of these data would have an impact on people's understanding of the study findings."[1]

Health technology assessment is a form of policy research that examines evidence of safety, efficacy, and cost effectiveness of a healthcare technology to provide guidance and recommendations to support decisions about treatment. Health technology assessment systematic reviews of clinical studies aim to include all relevant studies conducted on a particular topic and to provide an unbiased summary of their results, thus producing the best evidence on the benefits and harms of medical treatments. However, research has shown that the validity of systematic reviews can be affected by outcome reporting bias in the primary studies, which has been defined as selection (on the basis of the results) of a subset of the original variables recorded for inclusion in a study publication.[2]

The prevalence and impact of outcome reporting bias has recently been investigated in a large unselected cohort of Cochrane systematic reviews (ORBIT (Outcome Reporting Bias In Trials) I study).[3] This study, which focused on a single primary outcome for each review, found that over half the reviews (157/283, 55%) could not include data for the review primary outcome from all eligible studies. Additionally, interviews were conducted with trialists to understand the reasons for discrepancies between outcomes specified in the study protocol and those reported in the study publication.[1] The

Correspondence to: J J Kirkham jjk@liv.ac.uk

Extra material supplied by the author (see http://www.bmj.com/content/349/bmj.g6501?tab=related#datasupp)

Figure showing examples of specification of harm outcomes in systematic reviews
Table showing examples of specification of harm outcomes in systematic reviews
Table showing examples of study classifications for harm outcomes

main finding was that trialists had reported the outcomes in a biased way in over a quarter of the studies.

The ORBIT I study focused primarily on selective non-reporting bias for benefit outcomes. Empirical evidence suggests that the reporting of data on harms is likely to be less complete than that of efficacy measures.[4] There are also many studies that have previously measured poor adverse event reporting in primary studies and systematic reviews, and all concluded that the reporting of harm outcomes is problematic.[5-7] Interviews with trialists have revealed that in one instance important harm outcomes were subject to selective non-reporting because the trialists preferred to focus on the positive benefits of an intervention, while leaving "undesirable" data on harms unreported.[1]

In the current study, ORBIT II, we estimated the prevalence of selective non-reporting of outcomes in studies within a cohort of published meta-analyses, where the outcome was harm. We developed a new classification system for the assessment of selective non-reporting of harm outcomes, and we demonstrate the application of this classification in two cohorts of systematic reviews to characterise the reasons for the non-reporting of harms data.

Methods

Data source

We examined both an unselected cohort of new reviews from the Cochrane Library (Cochrane cohort) and a separate cohort of systematic reviews of adverse events (adverse event cohort). Each month the Cochrane Library indexes newly published reviews in each monthly issue as "new reviews." PS identified all new Cochrane reviews from six monthly issues (issue 9, 2012 to issue 2, 2013). For the adverse event cohort, we had access to a bibliographic list of potentially eligible systematic reviews of adverse effects that had previously been identified and evaluated in other publications.[8-10] In brief, two separate researchers (YKL and other colleagues) checked all records that were published between 2007 and 2011 in the Cochrane Database of Systematic Reviews and the Database of Abstracts of Reviews of Effects. A review was included in the bibliographic list if the primary aim was to evaluate adverse effects associated with any healthcare intervention (as stated in the objectives of the review by the author). We excluded reviews in which adverse effects were a secondary outcome or the objective was to evaluate the "effectiveness and safety" of an intervention. The two cohorts of reviews were therefore clearly distinguishable; the Cochrane cohort typically consisted of reviews that aimed to assess the beneficial and unintended effects of an intervention within the same reviews. In contrast, the adverse event cohort consisted of reviews that were designed to focus on harms, and as such would have used search strategies tailored towards identifying studies reporting on specific harms. We included reviews that included exclusively randomised controlled trials and a combination of randomised controlled trials and non-randomised studies. We excluded reviews that considered only non-randomised studies, methodological and diagnostic test accuracy studies, and overviews of reviews.

Assessment of systematic reviews

For each review in the cohorts, JJK or PS examined the section referring to types of outcome measures to determine whether the review specified no harm outcome, harm outcomes pooled together (for example, any adverse events or all adverse events), or a specific harm outcome (for example, gastrointestinal bleeding or heart failure). The lead reviewer was then contacted and asked to confirm our categorisation and to select the single most important harm outcome if multiple specific harms were listed. Where no contact could be established, a clinical professor in clinical pharmacology and co-convenor of the Cochrane Adverse Effects Methods Group (YKL) reviewed our categorisations and where necessary selected a single primary harm outcome from those listed. The supplementary figure and table provide examples of reviews where the review did not specifically evaluate harm outcomes, the single primary harm outcome was clear from the review, the single primary harm outcome had to be chosen by the reviewer or pharmacological expert, and only pooled harms were specified.

We then further scrutinised the reviews where a single primary harm outcome had been defined. Reviews that did not identify any studies (empty review) were not assessed further. Equally, reviews that specified no harm outcomes or reported only on pooled harm outcomes were excluded from further consideration because an assessment of outcome reporting bias in these situations would not be possible or feasible. Two investigators (PS and JJK) scrutinised each review to check whether all included studies fully reported the single primary harm outcome.

Reviews were also eligible for further assessment if any primary studies had been excluded from the review for the single reason of not reporting on any review outcomes of interest. This is because potential outcome reporting bias can occur when a study report does not mention or give results for particular outcomes; absence of data on adverse effects does not necessarily mean that such outcomes were not measured or analysed. We therefore checked the list of excluded studies within a systematic review and reasons for exclusion to see if a study had been inappropriately excluded because the study did not report on any of the relevant review outcomes. If the references to excluded studies of interest were not provided by the reviews (in the adverse event cohort), we contacted the lead authors for clarification. If these study reports were not available on request, then we excluded the review from further assessment because a complete assessment could not be performed.

We selected a sample of 50 eligible reviews from each cohort for further assessment. The 50 reviews were randomly sampled from the list of eligible reviews that contained no more than 20 studies. This restriction was applied because the ORBIT I study showed that review authors were less likely to respond if the burden of work was too great; this typically arose when a review included a large number of studies requiring assessment.

Classification of studies in the reviews

Full reporting

We considered full reporting to have taken place if outcome data on the single primary harm outcome from the primary studies had been included by reviewers in the review meta-analysis (or tabulated in full if a meta-analysis was not appropriate). Such outcome data may have been included in the study report or indirectly calculated from the results. For binary outcome data, we considered full reporting to have occurred if the sample size was reported in each group together with either the number of participants with the event for each group or the odds ratio or relative risk with precise P value or a measure of precision or variability (confidence interval or standard error). For continuous outcome data, we judged that full reporting had taken place if the study described group sample sizes and magnitude of the treatment effect (for example, group means or medians or mean differences), and an exact P value or a measure of precision or variability (confidence interval, standard deviation, or standard error). When data on the single primary

harm outcome were missing or incompletely described in the review, but detailed examination of the primary study showed that the single primary harm outcome had actually been fully reported, we classified this situation as full reporting. If data were available (but somehow omitted from the review), we provided feedback to review authors so that the data could be included in an update of the review.

Not reported or partially reported

We developed a classification system to assess whether a study had been excluded from a meta-analysis or the descriptive analysis because the data for the harm outcome were either not reported or only partially reported. The system was designed to make a judgment about the reason for missing data on harm outcomes and was refined over the initial few months of the study. When any amendments were made, we reviewed all previous classifications and adjusted them accordingly to ensure consistency of application. The categories reflect the stages of assessing whether a specific harm outcome was measured or compared, and finally the nature of the results presented (table 1⇓). We used the category letters (P-V) to distinguish them from the A-I classifications used for benefit outcomes in the ORBIT I study.[3] The classification system was designed to be used with both randomised controlled trials and non-randomised studies. A supplementary table provides examples and guidance of when to use each of the ORBIT II classifications.

In the context of harm outcomes, we awarded classifications for "high risk" outcome reporting bias when the specific harm had been measured but the data were presented or suppressed in a way that would mask the harm profile of particular interventions (including providing detail on the seriousness of the harms)—that is, P1, P2, R, and S classifications (table 1). The reasoning behind this particular definition of bias is that in all these situations we at least suspect that the harm outcome had been measured and thus the selective non-reporting or partial reporting could have been driven by a biased related reason (for example, higher frequency of harm in one treatment arm). Missing data as a result of a P1, P2, R, or S classification are also likely to have the biggest impact on the treatment effect in a meta-analysis of reviews.

Using all the identified publications for a study, one investigator (PS or JJK) classified any study that did not fully report the review single primary harm outcome of interest. This also involved evaluating excluded studies that had been selected for assessment. Review authors were also contacted to answer any queries and to provide their expert clinical judgments on the classifications for each study. For each classification, justification for the categorisation was recorded in prose to supplement the category code, including verbatim quotes from the study publication whenever possible. When the coauthors of the corresponding review were unable to assist with our assessments, an expert in adverse events (YKL) also completed the assessment. This approach ensured that each assessment had input from one clinical expert and one non-clinical methodological expert. Any complex cases were discussed with experts in trial design and management (DA, CG, or PRW) at study group meetings. Discrepancies were discussed until a final overall classification was agreed for each study and the justification for the classification documented in full. To check for consistency, at the end of the study JJK reviewed all agreed classifications, with the justification. This additional quality assurance step ensured that the same classification was awarded to studies that had missing or partially reported data on harms for the same reason. Supplementary figure A provides a summary of the study methods.

Results

Description of review cohorts

Cochrane cohort

The Cochrane Library published 243 new reviews in issue 9, 2012 to issue 2, 2013 (fig 1⇓). Specific harm outcomes were listed in 92 reviews, of which 16 identified a single primary harm outcome in the review text. For the remaining 76 reviews, lead reviewers, coauthors, or YKL selected a single primary harm outcome from those specific harms listed (which happened to be the first listed harm outcome in 55 (72%) of the reviews). In the 92 reviews that aimed to assess specific harms, the median number of studies per review was 7 (range 0-75, interquartile range 3-12). A total of 13 reviews did not require further assessment: eight did not identify any studies (empty review) and five fully reported the single review primary harm outcome for all eligible studies. Further assessment was required in the remaining 79/92 reviews (86%) as they did not include full data on the single review primary harm outcome from all eligible studies.

Adverse event cohort

A total of 234 reviews published between 2007 and 2011 were identified in the adverse event cohort (fig 2⇓). Specific harms were listed in all but four reviews, where only pooled harms were stated. A single primary harm outcome was clearly defined in the review text for 190 of the 230 reviews where specific harms were stated. The single primary harm outcome for the remaining 40 reviews was selected by lead reviewers, coauthors, or YKL (this was the first listed harm outcome in 36 (90%) reviews). In the 230 reviews specifying specific harms, the median number of studies per review was 18 (range 1-209, interquartile range 16-29). A total of 57 reviews did not require further assessment as they fully reported the single review primary harm outcome for all eligible studies. Further assessment was required in the remaining 173/230 reviews (76%) as they did not include full data on the single review primary harm outcome from all eligible studies. However, a complete assessment could only be done in about half of these reviews (n=86) because citations of studies that were excluded owing to "no relevant outcome data" were not provided in the reference list of the review and were not obtainable from the review authors.

Description of studies with missing single primary harm outcome in review

We assessed the 92 reviews in the Cochrane cohort and found that the review single primary harm outcome was partially reported or not reported for 705 (76%) of the 931 studies included (fig 3⇓). For the adverse event cohort, there were a total of 8837 studies across all 230 reviews where a single primary harm outcome was specified, 7720 of which were implicated in the 173 reviews where further outcome reporting bias assessment was required (fig 4⇓). Nearly half (47%, 4159/8837) of the studies did not report or had partially reported the single primary harm outcome; only 796 of these were included in the 86 reviews for which a complete assessment could be carried out. References were unavailable for 2337 studies that were excluded owing to "no relevant data," of which the design of the study was unknown in 1637 (fig 4).

Classification of studies

Across the sample of 50 reviews from each cohort, data on the single primary harm outcome were missing in 41% (486/1178)

of the studies (407 included studies, 79 excluded studies; 375 randomised controlled trials, 111 non-randomised studies, fig 5⇓).

Table 2⇓ shows the classification of the 486 studies where the single primary harm outcome data were either missing or partially reported in the review (170 studies from the Cochrane cohort and 316 from the adverse effect cohort). Nearly a fifth of assessed studies (19%; 93/486) provided full data on the single primary harm outcome of interest that was not included in the review (table 2). Forty eight studies reported actual event rates for each treatment arm or a suitable effect estimate with a corresponding measure of precision, 23 specified that there were no harms observed in the study, 21 specified there were no actual events in the study concerning the single primary harm outcome, and one reported the data in full in a way that was not suitable for inclusion in the review analysis.

In nearly a third of studies assessed (32%, 126/393) where a missing data classification was needed, it was clear that the single review primary harm outcome was measured (P, Q, or R classifications) but either not reported or partially reported. In over half of the studies (53%, 208/393), it was likely that the single review primary harm outcome was measured (S or T classification) but not mentioned specifically. Outcome reporting bias was suspected in 63% (248/393) of studies (P1, P2, R, and S classifications).

Discussion

The principle findings of the study showed a high number of reviews suspected of outcome reporting bias as a result of missing or partially reported harm related outcomes. In this study, an assessment of selective non-reporting of outcomes in a review was required if one or more studies eligible for inclusion in the review did not report data on the single review primary harm outcome. An assessment was needed in 86% of reviews in the Cochrane cohort and 76% of reviews in the adverse event cohort. The proportion of reviews requiring assessment for harm related outcomes was substantially higher than that found in the previous Outcome Reporting Bias In Trials (ORBIT) I study in which 55% (157/283) of Cochrane reviews did not include full data for the single review primary benefit outcome of interest from all eligible trials.[3] In the ORBIT I study, 31% of trials did not report on the primary benefit outcome; this compares with 76% of the studies not reporting on the single primary harm outcome from the Cochrane cohort and 47% from the adverse event cohort. In ORBIT I, 50% of trials (359/712) were under high suspicion of outcome reporting bias, compared with 63% (248/393) in this study (P1, P2, R, and S classifications). Readers of systematic reviews should be aware that outcome reporting bias could potentially have an important impact on the effect size estimates of adverse events.

Strengths and limitations of this study

The strengths of this study are that we evaluated two large cohorts of reviews. Most reviews in the Cochrane cohort quantified both the beneficial and the harmful effects of healthcare interventions within the same review. The adverse event cohort of reviews aimed specifically to synthesise studies reporting on specific harms, which may have been longer term harms. No reviews overlapped in the two cohorts (they stemmed from non-overlapping time periods), and none of the studies assessed for outcome reporting bias were included in multiple reviews. Since non-randomised studies are often needed to address questions about serious, rare, and long term harmful or unintended effects, our classification system was developed to encompass missing harms outcome data for both randomised controlled trials and non-randomised studies. Review authors or an expert in clinical pharmacology were involved in the assessments, and a senior investigator checked a textual justification for each classification. All assessments were therefore carried out by a clinical expert and a non-clinical methodologist. Importantly, more than two thirds of the reviews assessed in the Cochrane cohort had input on classification from the lead reviews. There was relatively less reviewer input for the adverse event cohort, which is not unexpected given that this cohort of reviews was slightly older.

In this study, we made a judgment as to whether the review single primary harm outcome was measured or not based on all the study references listed in the review. One way to improve this judgment would be to obtain study protocols of eligible studies to compare prespecified outcomes with those reported in the final study reports. The comparison between what was planned and assumed to be measured (in terms of outcome specification) and what was actually reported (or not reported) has the potential to simplify the classification of missing outcome data.

We suspect that there are three aspects of this study that may lead us to underestimate the extent of outcome reporting bias. To adopt a similar approach to ORBIT I, we chose to look at one primary review harm outcome for assessment from each review. When choosing a single primary harm outcome, it is possible that these were selected by the reviewers or pharmacological expert because it was the particular harm that was well recognised, important, serious, or common. This means that the harm outcomes evaluated in our study would have been less prone to outcome reporting bias. Hence our findings may actually turn out to be an underestimate of the greater scale of the problem. In practice, individual reviewers should complete the assessment for all specified outcomes; this was not feasible in our study because of the large cohort of reviews. Secondly, the absence of references to excluded studies meant that nearly half of the reviews in the adverse event cohort were excluded because it was not possible for us to complete a full assessment of outcome reporting bias. Although these exclusions were unavoidable, we acknowledge that there may be greater suspicion of outcome reporting bias in such reviews owing to these excluded studies not reporting on any relevant outcome data when compared with those reviews we were able to assess within the same adverse events cohort. Finally, owing to the high prevalence of reviews affected by missing single primary harm outcome data, we were only able to fully assess a sample of 50 reviews from each cohort of reviews. We do not suspect that our sampling strategy will have impacted importantly on the results of this study as the reviews were unselected in terms of the type of harm, intervention, or population. However, if there were more studies in any particular specialty, there might be a greater expectation that specific harms would have been measured, so the suspicion of outcome reporting bias would be higher in the reviews we did not look at..

Implications for reporting harms in studies of healthcare interventions

Many of the classifications for the partial or non-reporting of harms data found in this study relate to poor reporting practice. Better reporting of harms would provide timely and important information to guide clinicians and the public in making decisions about treatment not only at the individual study level but also through allowing harms to be properly investigated in systematic reviews and meta-analysis, which can have greater statistical power from pooled data. We appreciate that many

individual studies will be underpowered to detect significant differences between groups for harm outcomes. Nevertheless, it is important that these individual studies fully report on all harms data, including "zero events" and harms that are rare.

To help monitor the complete reporting of harms (for randomised controlled trials at least), researchers or independent boards (sponsors, research ethics committees, or regulatory authorities) must develop and implement plans to monitor and report data for the safety of participants. The monitoring plan should evaluate and report the incidence and severity of expected harms to confirm that they match with those expected at the initiation of the research. The incidence and causality of any unanticipated or unexpected harms should also be closely monitored and reported. Adherence to the CONSORT (consolidating standards of reporting trials) extension for harms [11] or STROBE (strengthening the reporting of observational studies in epidemiology)[12] for non-randomised studies could also greatly reduce the problem of poor reporting of harms. Nevertheless, a recent review evaluating studies that have examined the influence of the CONSORT for harms criteria has shown that adverse events are poorly defined, with six of the seven included studies demonstrating less than 50% adherence to the items on the checklist.[13] Journals should be more explicit in their recommendations and expectations regarding the endorsement of the CONSORT or STROBE statement and related extensions (for example, CONSORT extension for harms). Few journals recommend or require specific endorsement of the CONSORT for harms statement in their instructions for authors' section.[14]

Implications for systematic reviews

The reliability of systematic reviews can be improved if more attention is paid to specifying harm outcomes in a review. The Cochrane handbook specifically states "There should in general be no more than three primary outcomes and they should include at least one desirable and at least one undesirable outcome (to assess beneficial and adverse effects, respectively)."[15] Despite such guidance nearly a fifth (17%) of newly published reviews in the Cochrane cohort did not specify any harm outcomes, and 44% specified only pooled harms. This is only a marginal improvement on the 24% (18/78) of Cochrane reviews published in issue 1, 2005 that did not report on adverse events.[6] In addition, a recent study of 296 reviews identified from the Database of Abstracts of Reviews of Effects revealed that nearly a third of these reviews did not clearly define the adverse events reviewed.[16] It was not feasible to assess outcome reporting bias in reviews that dealt with only pooled harms because there would be no way to assess whether specific harms may have been excluded from the eligible studies when calculating "all harms." Moreover, we discovered that nearly a fifth of studies that underwent an assessment for outcome reporting bias had actually provided full data (which had somehow been omitted from the review) on the single primary harm outcome. We strongly recommend that authors of systematic reviews pay more attention to declaring the important harms for inclusion in the review at the protocol stage. Extra care is required during data extraction to ensure that reported outcomes data from included studies is not missed; we also believe that for complete transparency, data on zero events or no harms should also be reported in reviews. In this study, any missing data on harms from reviews that were found in study reports were fed back to all review authors.

Studies should not be excluded from reviews because of having" no relevant outcome data," as the outcome data may be missing as a direct result of selective outcome reporting.[3] Nearly half the reviews requiring assessment in the adverse event cohort could not be assessed because studies were excluded for this reason and no record was kept of the references to these excluded studies.

The classification system used in this study has been presented and applied during a workshop that we developed and delivered at international Cochrane colloquiums. The feedback from this workshop supported the practical application of our classification system, and many participants were able to relate their own experiences to the types of scenarios that are captured in the classification. Following the application of the classification system, the Cochrane risk of bias tool is currently being updated to include the assessment of bias in both randomised controlled trials and non-randomised studies. The proposed new structure of the risk of bias tool considers selective outcome reporting as being analogous to publication bias (non-reporting of whole studies). It is planned that this form of bias will be appraised outside the risk of bias tool (for example, as part of the GRADE assessment in the summary of findings tables).[17] Our classification system does not confirm bias but will help reviewers gain a better understanding of the reasons for the lack of detail on harm reporting and which mechanisms may be at high risk of bias. The implementation, writing of guidelines, and Cochrane handbook chapters will include guidance from both ORBIT I and ORBIT II; this should raise the awareness of this problem for both benefit and harm outcomes among the community of systematic review authors. This in turn should improve the ability of decision makers to make informed choices that consider both the benefits and the harms of an intervention in an unbiased way, ultimately improving patient safety.

Future research

The poor reporting of harms data in studies and systematic reviews has implications for clinicians and patients because there are difficulties in judging the benefit-risk trade-off when much of the harms data is inadequately reported or not reported at all. However, the mechanism for bias in harms reporting remains unclear, although we are aware of a few potential scenarios. This includes the conscious desire to avoid publishing data that are unfavourable to a particular intervention. The opposite is true if study authors make their own personal judgment that particular adverse events are not serious or unimportant and not significant, and therefore do not merit being reported.

Owing to problems with recall bias in the earlier interview study[1] and because little emphasis was placed on harm outcomes, we plan in future to interview trialists about differences between outcomes specified in trial protocols and the trial report during the peer review process to better understand mechanisms for outcome reporting bias across both benefit and harm outcomes. In partnership with *The BMJ*, we are conducting a pilot study to determine the feasibility of carrying out such interviews. We also plan to write a separate tutorial paper for assessing outcome reporting bias in all benefit and harm outcomes for a single review using the methods from ORBIT I and ORBIT II.

Conclusions

Our investigation found that many reviews were affected by missing data on harms from at least one eligible study. We suspected high risk of outcome reporting bias as the cause of the missing data in over half the studies assessed. There is a clear need to raise the awareness of both the existence and the potential impact of bias when study authors measure harm

outcomes and then to choose to either selectively not report the findings or present the results in a way that cannot be reliably used in a systematic review.

We thank the reviewers whose collaboration made this research possible. Their input includes defining the review primary harm outcome of interest, forwarding study reports from their reviews, and establishing the outcome reporting bias classification for particular studies within their reviews.

Contributors: DGA, CG, JJK, and PRW obtained the study funding and designed the study. JJK, YL, and PS acquired the data. All authors analysed and interpreted the data. JJK and PS prepared the initial manuscript and performed the statistical analysis. DGA, CG, JJK, YL, and PRW supervised the study and critically revised the manuscript. All authors commented on the final manuscript before submission. PS was responsible for administration, technical, and material support. JJK is the guarantor.

Funding: The Outcome Reporting Bias In Trials (ORBIT II) project was funded by the Medical Research Council (grant No MR/J004855/1). DGA is supported by a Cancer Research UL programme grant (C5529). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of this manuscript.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisation that might have an interest in the submitted work in the previous three years; YL is co-convenor of the Cochrane Adverse Effects Methods Group; however, the authors have no other relationships or activities that could appear to have influenced the submitted work.

Ethical approval: Not required.

Data sharing: The data from this study are available from the corresponding author (jjk@liv.ac.uk). For each of the two cohorts of reviews, details on the eligibility of reviews for further assessment are available; this includes the selection of the single primary harm outcome and summaries of the number of studies reporting or not reporting on the single review primary harm outcome from each review. For the sample of reviews where a full outcome reporting bias assessment was undertaken, final study outcome reporting bias classifications and justifications for classifications are all available.

Transparency: The manuscript's guarantor (JJK) affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

1 Smyth R, Kirkham JJ, Jacoby A, Altman DG, Gamble C, Williamson PR. Frequency and reasons for outcome reporting bias in clinical trials: interviews with trialists. *BMJ* 2011;342:c7153.
2 Hutton JL, Williamson PR. Bias in meta-analysis due to outcome variable selection within studies. *Appl Stat* 2000;49:359-70.
3 Kirkham JJ, Dwan KM, Altman DG, Gamble C, Dodd S, Smyth R, et al. The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *BMJ* 2010;340:c365.
4 Chan A-W, Krleza-Jeric K, Schmid I, Altman DG. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
5 Jonsson U, Alaie I, Parling T, Arnberg FK. Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioural disorders: a review of current practice. *Contemp Clin Trials* 2014;38:1-8.
6 Hopewell S, Wolfenden L, Clarke M. Reporting of adverse events in systematic reviews can be improved: survey results. *J Clin Epidemiol* 2008;61:597-602.
7 Papanikolaou PN, Ioannidis JPA. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med* 2004;117:582-9.
8 Golder S, Loke Y, Zorzela L. Some improvements are apparent in identifying adverse effects in systematic reviews from 1994 to 2011. *J Clin Epidemiol* 2013;66:253-60.
9 Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ* 2014;348:f7668.
10 Golder S, Loke YK, Zorzela L. Comparison of search strategies in systematic reviews of adverse effects to other systematic reviews. *Health Info Libr J* 2014;31:92-105.
11 Ioannidis JP, Evans SJ, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141:781-8.
12 Von Elm E, Altman D, Egger M, Pocock SJ, Gotzche PC, Vandenbroucke JP, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med* 2007;147:573-7.
13 Hodkinson A, Kirkham JJ, Tudur-Smith C, Gamble C. Reporting of harms data in RCTs: a systematic review of empirical assessments against the CONSORT harms extension. *BMJ Open* 2013;3:e003436.
14 Hopewell S, Altman D, Moher D, Schulz KF. Endorsement of the CONSORT statement by high impact factor medical journals: a survey of journal editors and journal 'Instructions to Authors'. *Trials* 2008;9:20.
15 Higgins JPT, Green S, eds. Cochrane handbook for systematic reviews of interventions, version 5.1.0 [updated March 2011]. Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org.
16 Zorzela L, Golder S, Liu Y, Pilkington K, Hartling L, Joffe A, et al. Quality of reporting in systematic reviews of adverse events: systematic review. *BMJ* 2014;348:f7668.
17 Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidenced publication bias. *J Clin Epidem* 2011;64:1277-82.

Accepted: 16 October 2014

Cite this as: *BMJ* 2014;349:g6501

What is already known on this topic

Empirical evidence suggests that the reporting of data on harms is likely to be less complete than that of benefit outcomes

Outcome reporting bias has previously been identified as a threat to evidence based medicine

Although previous research involving benefit outcomes has suggested that outcome reporting bias was suspected in at least one trial in over a third of reviews, little is known about the prevalence of outcome reporting bias in harm outcomes

What this study adds

Outcome reporting bias for harms was evident in nearly two thirds of all primary studies included in systematic reviews

The reliability of systematic reviews can be improved if more attention is paid to specifying harm outcomes in a review

Studies should not be excluded from reviews because of having" no relevant outcome data," as the outcome data may be missing as a direct result of selective outcome reporting

Tables

Table 1| **ORBIT II classification system**

Classification	Description	Level of reporting	Risk of bias*
Explicit specific harm outcome			
Measured and compared across treatment groups:			
P1	States outcome analysed but reported only that P>0.05	Partial	High risk
P2	States outcome analysed but reported only that P<0.05	Partial	High risk
P3	Insufficient reporting for meta-analysis or full tabulation	Partial	Low risk
Measured but not compared across treatment groups:			
Q	Clear that outcome was measured and clear outcome was not compared	NA	No risk
Measured, not clear whether compared or not across treatment groups†			
R1	Clear that outcome was measured but no results reported	None	High risk
R2	Result reported globally across all groups	None	High risk
R3	Result reported from some groups only	None	High risk
Specific harm outcome not explicitly mentioned			
Clinical judgment says likely measured and likely compared across treatment groups:			
S1	Only pooled adverse events reported (could include specific harm outcome)	None	High risk
S2	No harms mentioned or reported	None	High risk
Clinical judgment says likely measured but no events:			
T1	Specific harm not mentioned but all other specific harms fully reported	None	Low risk
T2	No description of specific harms	None	Low risk
Specific harm outcome not explicitly mentioned, clinical judgment says unlikely measured			
U	No harms mentioned or reported	None	Low risk
Explicit the specific harm outcome was not measured			
V	Report clearly specifies that data on specific harm of interest was not measured	NA	No risk

NA=not applicable (clear that outcome was not going to be compared).

*Bias would occur if specific harm had been measured, but data were presented or suppressed in a way that would mask the harm profile of particular interventions.

†Clinical judgment says likely measured and likely compared across treatment groups.

Table 2| **Studies assessed for outcome reporting bias**

	Cochrane cohorts		Adverse event cohorts		
Classification	Randomised controlled trial (%*)	Non-randomised studies (%†)	Randomised controlled trial (%‡)	Non-randomised studies (%§)	Total No of studies (%¶)
P1	6 (4.8)	0 (0.0)	7 (4.1)	5 (5.6)	18 (4.6)
P2	0 (0.0)	0 (0.0)	0 (0.0)	4 (4.4)	4 (1.0)
P3	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.1)	1 (0.3)
Q	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0)
R1	5 (4.0)	0 (0.0)	21 (12.2)	16 (17.8)	42 (10.7)
R2	7 (5.6)	0 (0.0)	25 (14.5)	13 (14.4)	45 (11.5)
R3	7 (5.6)	0 (0.0)	5 (2.9)	4 (4.4)	16 (4.1)
S1	2 (1.6)	0 (0.0)	12 (7.0)	4 (4.4)	18 (4.6)
S2	31 (24.8)	0 (0.0)	50 (29.1)	24 (26.7)	105 (26.7)
T1	5 (4.0)	3 (50.0)	3 (1.7)	2 (2.2)	13 (3.3)
T2	35 (28.0)	0 (0.0)	36 (20.9)	1 (1.1)	72 (18.3)
U	27 (21.6)	3 (50.0)	12 (7.0)	13 (14.4)	55 (14.0)
V	0 (0)	0 (0.0)	1 (0.6)	3 (3.3	4 (1.0)
Total	125	6	172	90	393
Full reporting**	38	1	40	14	93
Overall total	163	7	212	104	486

*Calculated as percentage of total number of randomised controlled trials in Cochrane cohort (excluding full reporting) (n=125).

†Calculated as percentage of total number of non-randomised studies in Cochrane cohort (excluding full reporting) (n=6).

‡Calculated as percentage of total number of randomised controlled trials in adverse events cohort (excluding full reporting) (n=172).

§Calculated as percentage of total number of non-randomised studies in adverse events cohort (excluding full reporting) (n=90).

¶Calculated as percentage of total number of classifications (excluding full reporting) (n=393).

**Review primary harm outcome data reported in full in study report but not in review.

Figures

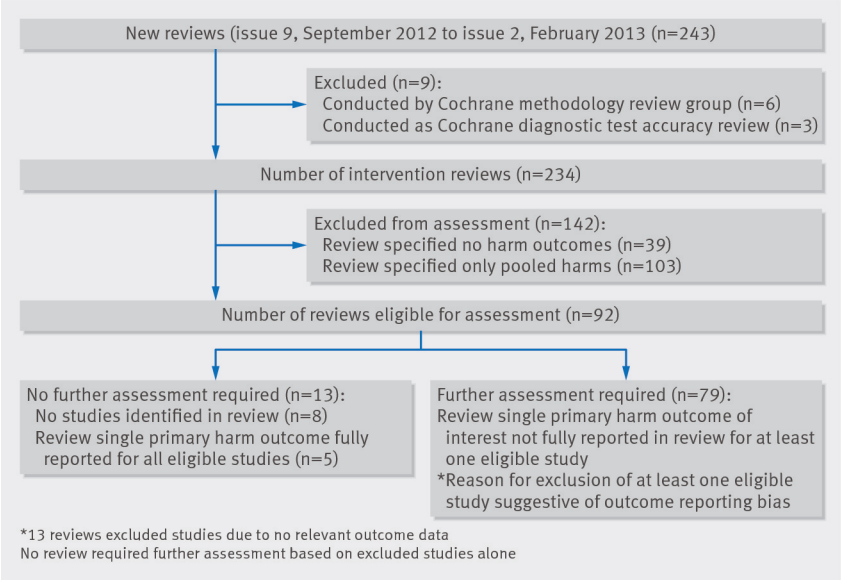

Fig 1 Flow diagram for Cochrane cohort

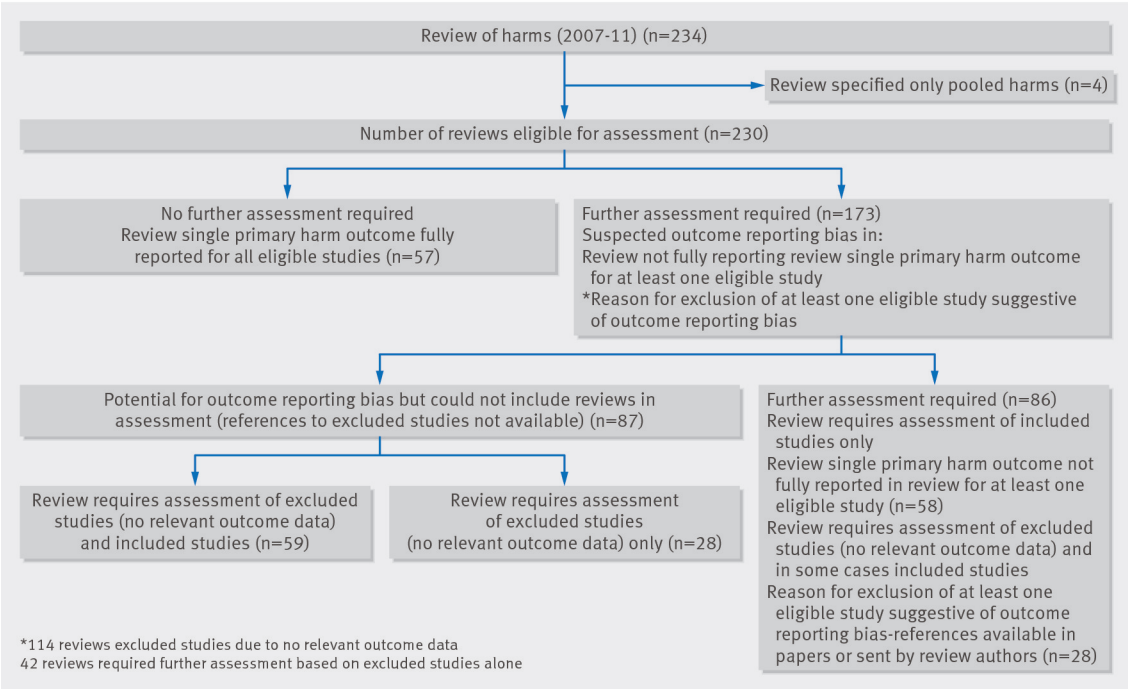

Fig 2 Flow diagram for adverse event cohort

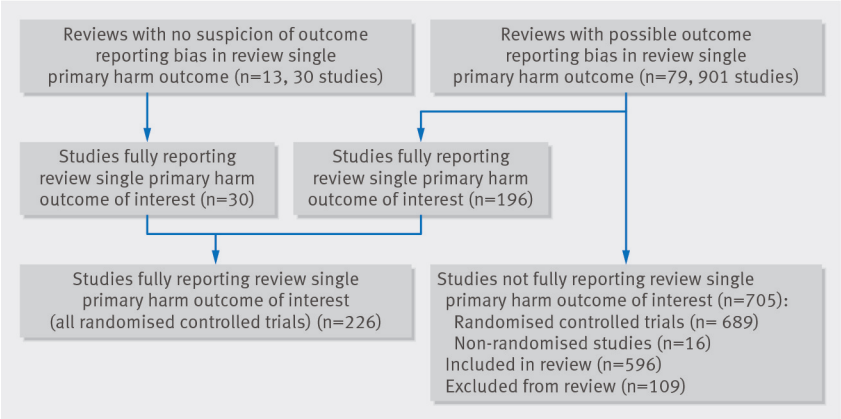

Fig 3 Assessment of studies within reviews (Cochrane cohort)

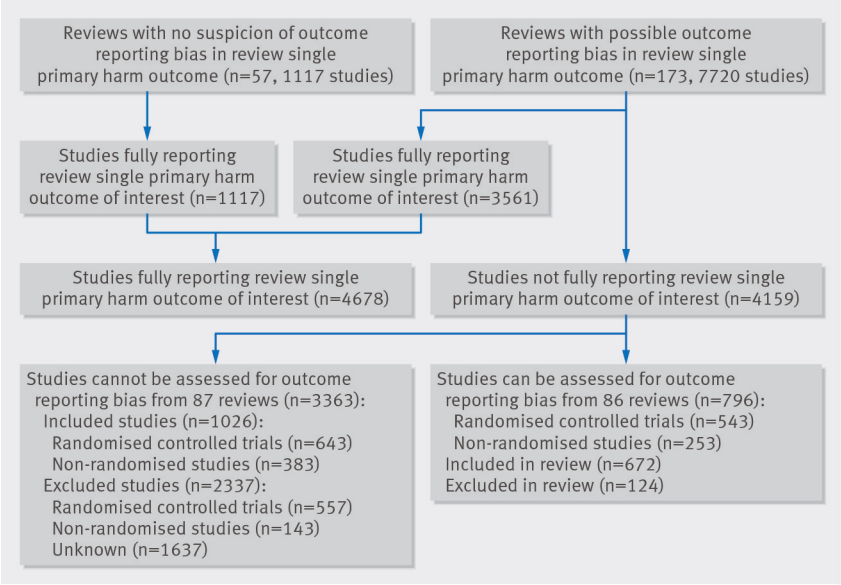

Fig 4 Assessment of studies within reviews (adverse event cohort)

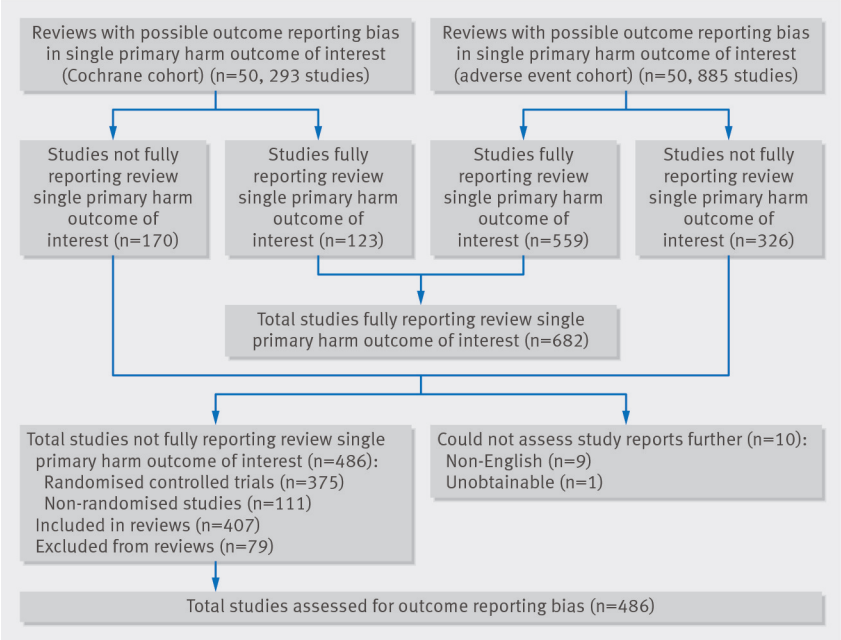

Fig 5 Assessment of studies from sample of 50 reviews from each cohort